2008 年度卒業論文

インターネット定点観測に基づくマルウェア平均感染の推定

研究指導　菊池浩明　教授

東海大学　電子情報学部　情報メディア学科

5ADM1118　松尾　峻治

目次

第1章 はじめに

セキュリティソフトをインストールされているPCがマルウェアに感染した場合、セキュリティソフトによって検出され削除される。一週間から一ヶ月に一回の割合でシステムをスキャンして、マルウェアを検出し、他のホストに広がるのを防ぎ、また、パターンマッチング用のシグネチャファイルの更新も頻繁に行い、すぐマルウェアによる被害が広がらないようにしている。しかし、実際はすぐ検出されず、数日経ってから検出されため、感染を防げないときがある。これは日々新しいマルウェアが開発されているのが一因である。IT技術の進化の一方で、IT技術を悪用して、マルウェアを作成する人たちも出てきているためである。マルウェアの対策を講じるのに十分な精度の解析をするには、長期的な観測が必要となるが、しかし、長期的な観測すると不正ホストも変化してしまい、解析の精度が落ちる可能性がある。このため、マルウェアのスキャンを観測する場合、最適な期間がどのくらいかわからない。また、同じホストが異なる時期に、複数回感染する可能性もある。現在セキュリティソフトの定義ファイルはマルウェアにすぐ対応できるように常に更新され続けているが、ネットワーク上では不正スキャンパケットが飛び交っており、マルウェアがすぐには駆除されていないことがわかる。では、過去マルウェアの感染期間は32日[1]と発表されたが、現在は32日であるのだろうか？本研究は2004年から2008年までのマルウェア感染期間の推移を求めることを目的とする。

第2章 研究目的

本研究はインターネット定点観測JPCERT/CCのISDAS[2]のログデータを使用している。過去32日[1]という感染期間がでた、これは約50台あるセンサのうち、不正パケットがよく観測された12台のセンサのログデータを使用しており、その12台中6台で観測された不正ホストを対象に実験が行われている。しかし、それではセンサ全部を使っている分けではなく、インターネット全体からの結果とは言いにくい。また、セキュリティソフトなどは一定期間でシグネチャファイルの更新をする傾向にあるため、マルウェアの感染期間は変動する。過去32日[1]を出したログデータは2004年から一年間のデータであるため、現在の感染期間は変化している可能性がある。そこで本研究はセンサ50台のログデータを使用して、2004年9月から半年毎の感染期間の推移を求めること目的とする。

第3章 感染期間の算出方法

3.1 視覚的感染期間算出方法

50 台のセンサの中から 12 台に注目し、そのうち6台で観測された 2007 年 11 月から 2008 年 4 月までのログデータを使い、不正ホスト約 1000 のログデータからパケットの動きをグラフにして感染期間を求める。その例を図 1 にて説明する。

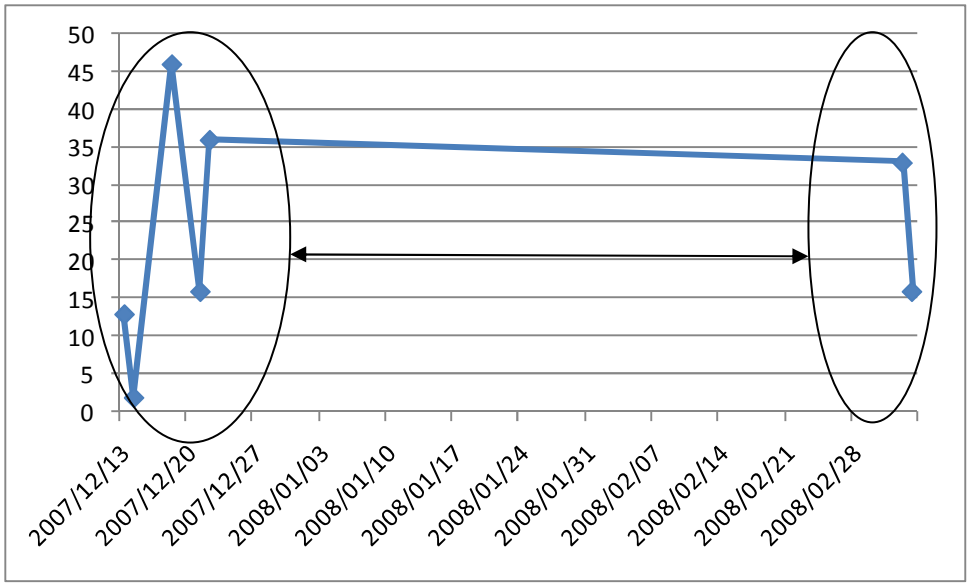

図 1:センサ番号とパケット到着間隔のグラフ

図 1 では X 軸をパケット到着間隔、Y 軸をセンサ番号としている。左右の丸の中それぞれにパケット 5 個、2個と飛んでいるがその丸の間の期間は何もパケットがない。丸の中のパケットの到着間隔と比べても明らかである。グラフから視覚的にホストの状態を判断する。この場合、それぞれ違うマルウェアに感染したと考え、丸のなかで最初と最後のパケットまで期間を感染期間と考える。この例では2つの感染期間があることがわかる。

3.2 適応閾値による算出方法

インターネット上のパケットを観測する方法に定点観測システムがある。本研究では、JPCERT/CC が運用しいている定点観測システム ISDAS[2]のセンンサで観測されたデータを元に解析した。センサとは、不正ホストとからのスキャンを観測する正規ホストである。[1]では、パケットの到着間隔は不正ホスト毎に異なり、感染・駆除のサイクルを決める閾値Tは一意にきまらないが、その到着感覚はポアソン分布に従うことが示されている。そこで本研究においてもパケット到着間隔はポアソン分布に従うと仮定し、期間tでパケットが届かない確率を同定する。λはパケットの到着率とし、$\lambda = c/d_0$と定義する。cは各不正ホストの総ケット数とし、d_0は最初と最後のパケットの観測時間差である。この時全パケット到着しない確率が 1%となる閾値は $T=-\ln(0.01)/\lambda$ で求められる。

第4章 調査報告

(ア) 視覚的感染期間の算出を行った結果

2007 年 11 月から 2008 年4月までのデータの中のセンサで多くパケットが観された12台中で、そのうちの 6 台で観測できた 1024 の不正ホストを対象に行った結果、平均感染期間は次のようになった。

表1:視覚的感染期間の算出結果

ラウンド数	ビジット数	感染期間	パケット数
1.686364	4.255378	18.04622	17.28242

これより 視覚的判断では、[1]で求められた感染期間 32 日より短くなっているのがわかる。

(イ) 適応閾値の算出をおこなった結果

2004 年 9 月 1 日～2008 年 4 月 30 日における独立したセンサ 50 台で観測されたデータを半年毎に分割(P_1～P_7)し、マルウェアの感染期間を同定した。この結果を表 2 に示す。ここで U をユニークホスト数、d を平均感染期間, σ を平均感染期間の標準偏差とする。また、感染機関の推移を図 2 に示す。

表 2:各観測期間の平均感染期間

	U_i	d_i	σ_i
P_1	1535990	11.52	25.37
P_2	1503238	11.34	26.14
P_3	950645	10.38	25.19
P_4	733038	8.20	21.54
P_5	790350	6.03	18.13
P_6	727947	5.97	17.96
P_7	552289	6.23	18.11

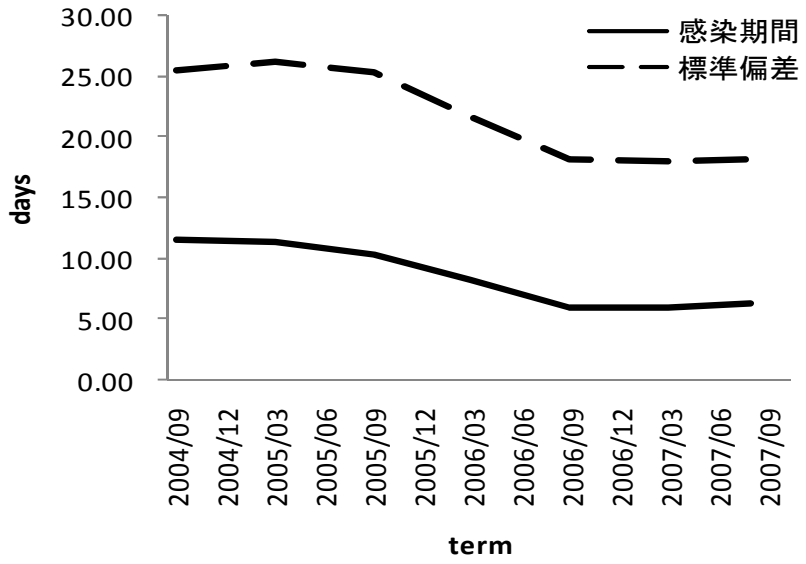

図 2:平均感染期間の推移介

表 1 の P_1 を見ると感染期間は 11.52 日となっており、[1]ので書かれている感染期間 32 日と同じログデータを使ってるにもかかわらず大幅に違う。これはセンサ 50 台のうち 12 台をランダムで選んで、その中の 6 台で観測された不正ホストから算出したためである。今回行った算出ではすべてのセンサから算出を行っているため、感染期間が1日となるものが多くなり、結果 11.52 日という期間となった。

また表 1 から、ユニークホスト数は P_1 の比べ P_7 の時期には 35%にまで下がっているが、感染期間は 54%、感染期間の標準偏差は 71%となっている。ユニークホストの減少に比べ、標準偏差の減少はあまり感じられない。標準偏差とは基本ばらつきを見るものである。つまり、まだマルウェアの種類によっては未発見のままであったり、対策がおくれていたりと、感染期間が長くなっている可能性がある。

第5章 考察

今回約 4 年間のログデータを半年ごとに分けて感染期間を求めた結果、感染期間が年々減少するのが見えた。このことより、現在マルウェア対策が進んでおり、マルウェアに感染したホスト減るのがわかる。実際ユニーク不正ホスト数は減少していた。

第6章　結論と今後の課題

結論として感染期間は減り、セキュリティが強くなってきていることがわかった。しかし、今回はあくまでログデータを半年毎にわけて、7個のデータからの予想のみで考えており、離散的なデータでしかない。今後の課題として、より連続的なデータを得る必要がある。また、連続的なデータをえることにより、1次以上の近似の求めることにより、感染期間の予測が立てられるのではないかと考えられる。

付録 1 プログラム仕様

1.1)ログデータについて

1.1.1 今回使用したログデータはテキストで 100MB 以上がほとんどであり、通常の Java のヒープ領域ではたりない。コマンドによってヒープ領域は確保できるが、Javaの処理時間が大変かかる。そのためログデータをすべて読み込むことはあまりお勧めできない方法であった。

1.1.2 そこで、プログラム上で工夫する前に、ログデータのほうで何とかならないかと考えた。通常ログデータは日付順にソートされている。これにより 1 つの不正ホスト感染期間を求めるには最後まで読み込まないと不正ホストのすべてのパケットを見ることができない。

1.1.3 そこで、ログデータの文字列を不正ホスト、日時という順に書き換え、ソートすることにより、ログデータをすべて読み込まなくても Java で動かせるようにした。これにより処理時間も 4 日かかる処理が 10 分ぐらいになったのは自分自身でも驚いた。

1.1.4 また、プライベートIPから飛んでくるものがあり、これはIPを偽装していると考え、観測対象から外している。

1.2)算出プログラムについて

1.2.1 ログデータを読みこんだ後、パケット間隔、ユニークセンサ数、ラウンド数を求めた。パケット間隔は今回日数で計算して、時間単位ではやらなかった。

1.2.2 何度か出力結果を見た結果、感染期間 1 日でパケットが 1 つしか飛んでない不正ホストが多いことがわかった。しかしこの場合感染しているとは考えられない。よって今回のケースは感染日数 0 日という形で数えている。

付録 2　ログデータの詳細

期間毎の詳細

1)　不正ホストの IP アドレス第一オクテット

今回使用したログデータの各期間の第一オクテットの統計を表 3、図 3 に示す。

表 3:第一オクテットのグラフ

IP_第一オクテット	Span1	Span2	Span3	Span4	Span5	Span6	Span7
0	0	2	60	33	2	0	0
1	2	0	95	134	1	2	1
2	1	0	84	126	2	0	0
3	3	1	297	532	13	0	0
4	29680	5850	2440	2524	1554	1789	1104
5	3	2	100	217	286	332	177
6	2	0	99	229	250	406	217
7	1	3	104	197	251	356	196
8	2373	18	318	719	797	871	489
9	1	0	92	239	279	350	230
10	0	0	0	0	0	0	0
11	2	6	142	237	288	399	197
12	8185	2155	1624	1935	2043	2175	1521
13	3	2	113	280	311	409	221
14	2	1	112	253	290	378	209
15	375	35	410	901	798	893	521
16	119	12	433	888	801	846	516
17	13	35	390	883	775	845	552
18	101	11	377	867	747	900	483
19	0	0	118	256	287	407	219
20	1	2	194	443	430	484	261
21	5	5	110	263	292	366	192
22	2	4	107	264	264	402	191
23	0	1	114	242	254	348	162
24	27804	8871	7568	11510	12508	8524	20291
25	3	1	113	256	270	401	198

26	0	0	115	261	262	410	191
27	3	0	129	241	233	397	168
28	2	1	116	244	251	356	216
29	0	0	115	275	262	410	191
30	0	2	113	276	244	395	193
31	0	0	122	256	216	354	184
32	106	38	381	855	737	928	452
33	9	2	388	922	813	904	502
34	1	4	98	298	275	378	208
35	126	30	361	924	775	862	478
36	0	0	117	259	240	328	199
37	0	3	110	241	266	347	213
38	315	96	485	1017	1005	1111	680
39	3	1	107	268	255	345	192
40	3	4	157	361	365	500	246
41	1	3	129	291	558	888	661
42	3	1	135	278	254	368	170
43	260	174	5781	4610	3552	2432	1238
44	11	4	344	844	734	872	483
45	5	7	351	513	697	618	350
46	2	1	141	242	291	393	188
47	2	5	192	370	404	452	268
48	0	2	104	254	303	418	213
49	3	1	91	218	291	376	213
50	2	4	116	255	233	327	197
51	2	2	150	241	308	454	177
52	3	1	110	240	283	407	210
53	4	8	361	813	785	896	489
54	0	1	112	259	288	401	180
55	5	7	380	820	727	899	480
56	1	2	143	287	280	401	183
57	66	18	400	851	791	859	504
58	152	20521	45859	35931	42487	26591	20312
59	2541	7091	9779	23860	35930	30577	22738
60	121380	67832	34899	21363	37679	13894	31920

61	955918	254496	153290	96422	67769	42526	34276
62	20280	7941	3519	2733	5675	5063	3106
63	9359	3044	1865	2025	2271	2345	1493
64	13981	4152	2749	3201	3667	4091	2800
65	19437	5311	5010	14568	9156	3285	2147
66	22568	6679	5072	5795	5810	5625	3595
67	16617	5724	3657	3430	3033	3137	2055
68	24145	7757	5937	5852	4552	4813	3039
69	17497	7442	5279	6676	4922	4350	2746
70	8488	5186	5514	6410	3840	3479	2231
71	357	6966	7946	8158	4793	4624	3151
72	143	401	1036	2652	3786	4519	2958
73	1	0	291	678	650	842	486
74	1	0	144	1367	1414	2538	1951
75	5	1	123	1541	1864	2344	1754
76	0	2	104	293	783	1759	1706
77	1	1	128	244	435	1514	1727
78	0	0	105	245	246	832	1603
79	0	3	133	249	288	932	1616
80	31585	10525	5314	3464	6058	5394	3309
81	46082	15249	5824	4123	6990	6133	3716
82	32254	13850	7462	5130	8105	7242	4480
83	24726	13738	7859	5574	16424	11254	5770
84	8405	11858	8708	5729	8241	6918	4647
85	3235	6398	4826	3966	4351	3746	2846
86	0	601	2123	2569	2927	3019	2094
87	2	71	1686	2430	6609	6659	4394
88	0	2	829	2440	6851	6357	4132
89	2	1	202	1093	6504	7865	5032
90	1	7	107	356	900	1604	1412
91	2	0	128	264	1009	1839	1678
92	4	2	128	254	265	382	555
93	1	2	129	253	245	366	203
94	0	0	138	235	242	322	191
95	1	1	94	263	261	346	202

96	0	2	118	227	304	476	420
97	1	0	117	233	279	478	408
98	2	4	104	242	294	477	456
99	3	4	144	242	281	480	563
100	3	2	111	254	241	367	186
101	2	2	112	264	267	366	203
102	0	0	115	248	267	327	197
103	2	0	115	239	263	345	213
104	3	1	114	248	246	382	198
105	2	3	110	239	258	370	203
106	7	1	112	237	246	350	198
107	1	2	116	251	251	330	192
108	2	0	141	245	234	338	190
109	0	2	123	250	271	350	206
110	1	2	131	255	240	337	197
111	3	2	135	254	249	367	218
112	1	2	131	231	241	346	208
113	1	1	127	265	263	361	199
114	0	0	115	226	272	375	215
115	0	2	122	237	271	324	211
116	3	1	121	262	252	896	1878
117	1	1	112	248	260	472	639
118	3	2	126	246	252	405	1015
119	1	0	126	238	248	389	228
120	1	3	126	252	270	366	200
121	1	1	125	611	18029	43341	31173
122	0	1	122	315	1994	7147	9058
123	2	0	122	288	503	2113	4018
124	0	0	177	4706	16802	22850	18131
125	0	13	5727	13200	21656	18723	16725
126	2	18	214	360	841	979	643
127	0	0	0	0	0	0	0
128	866	302	533	858	907	1027	582
129	1040	506	661	997	906	962	574
130	864	305	530	977	911	1093	617

131	355	106	432	809	837	870	506
132	192	86	435	834	786	870	477
133	71502	2111	1271	1184	1273	1135	773
134	318	129	425	724	697	738	477
135	32	34	262	531	509	621	357
136	72	37	243	474	511	559	289
137	418	168	450	787	702	829	471
138	1650	402	518	745	707	772	465
139	391	131	351	636	613	625	381
140	1202	461	668	856	827	991	665
141	3639	899	731	976	897	910	571
142	835	467	763	955	777	1037	629
143	289	158	418	725	654	761	509
144	941	533	521	789	748	897	429
145	675	326	421	701	676	815	497
146	439	103	315	632	644	733	394
147	402	172	418	702	649	755	419
148	2694	738	653	754	815	882	519
149	848	125	347	624	740	776	433
150	501	244	401	658	759	814	434
151	4056	1007	959	1109	974	1121	790
152	693	93	352	694	632	749	413
153	102	7	156	343	400	477	228
154	56	19	180	353	389	544	290
155	638	295	461	780	675	721	411
156	302	74	259	514	422	558	324
157	279	128	384	658	641	749	436
158	2192	63	330	689	579	720	386
159	457	194	386	637	600	700	415
160	235	120	311	613	611	709	385
161	311	123	316	587	554	642	398
162	1217	338	330	550	484	586	311
163	780	302	489	713	750	808	499
164	795	279	429	665	707	786	471
165	103943	12091	6167	1247	706	876	434

166	332	301	365	674	631	704	389
167	284	71	282	606	547	643	330
168	1383	359	565	857	769	845	459
169	1512	111	255	476	529	650	375
170	106	54	287	577	523	708	329
171	43	13	157	389	321	497	249
172	13067	4371	2658	2624	1577	2024	840
173	7	6	128	242	272	374	202
174	4	0	145	265	245	305	202
175	1	1	134	289	242	355	167
176	5	2	121	231	228	379	181
177	17	0	124	252	262	363	200
178	1	1	115	258	249	336	209
179	2	1	136	254	288	340	201
180	5	1	128	238	262	347	213
181	4	1	117	255	262	334	201
182	2	1	105	249	244	329	215
183	2	1	106	237	246	350	185
184	2	1	121	245	264	339	201
185	0	1	112	252	245	393	218
186	2	0	111	246	296	380	189
187	0	1	126	243	257	341	188
188	5	4	130	267	283	408	207
189	0	1	125	298	1378	2325	2264
190	1	3	243	712	1271	2331	1884
191	9	5	121	227	299	406	198
192	1503	281	521	652	715	872	518
193	4047	1442	1259	1380	2089	2128	1418
194	3483	1223	1091	1242	1256	1320	858
195	9359	3193	2297	1800	3314	3403	1845
196	999	971	1608	1816	1095	1235	691
197	4	3	105	222	239	381	210
198	1427	427	558	807	793	890	556
199	1440	399	545	714	799	904	583
200	37892	16278	9017	6770	12284	11798	5813

201	16191	12084	8552	6997	14737	13414	6247
202	55724	33886	18322	10874	12743	12995	10634
203	390989	48611	39982	22062	14427	13477	10351
204	2564	688	819	1245	1498	1468	959
205	2265	485	591	999	910	1092	677
206	3355	993	940	1402	1232	1346	777
207	6170	1990	1672	1926	2585	2511	1710
208	4224	1233	1161	1554	1762	2186	1357
209	10073	2867	1861	2019	2586	2813	1933
210	277238	108924	49712	22041	18971	19496	17015
211	76361	24392	27074	25281	31405	27465	18607
212	13871	5172	2771	2412	3836	3894	2435
213	25125	8293	4116	2896	5420	5295	3238
214	3	6	368	833	763	856	482
215	6	3	242	530	525	657	339
216	12929	4591	2868	2811	3797	3783	2596
217	31000	9759	3466	2674	4561	3918	2398
218	172635	95688	66109	37258	30168	19933	14372
219	1031451	171366	81952	54866	42922	51411	42752
220	907333	196819	80039	29890	31694	30877	19673
221	249445	132118	61473	46179	43861	23692	17146
222	821334	89698	74064	51218	34554	29666	14419
223	3	1	121	164	2	25	0
224	0	0	0	0	0	0	0
225	0	0	0	0	0	0	0
226	0	0	0	0	0	0	0
227	0	0	0	0	0	0	0
228	0	0	0	0	0	0	0
229	0	0	0	0	0	0	0
230	0	0	0	0	0	0	0
231	0	0	0	0	0	0	0
232	0	0	0	0	0	0	0
233	0	0	0	0	0	0	0
234	0	0	0	0	0	0	0
235	0	0	0	0	0	0	0

236	0	0	0	0	0	0	0
237	0	0	0	0	0	0	0
238	0	0	0	0	0	0	0
239	0	0	0	0	0	0	0
240	0	0	0	0	0	0	0
241	0	0	0	0	0	0	0
242	0	0	0	0	0	0	0
243	0	0	0	0	0	0	0
244	0	0	0	0	0	0	0
245	0	0	0	0	0	0	0
246	0	0	0	0	0	0	0
247	0	0	0	0	0	0	0
248	0	0	0	0	0	0	0
249	0	0	0	0	0	0	0
250	0	0	0	0	0	0	0
251	0	0	0	0	0	0	0
252	0	0	0	0	0	0	0
253	0	0	0	0	0	0	0
254	0	0	0	0	0	0	0
255	0	0	0	0	0	0	0

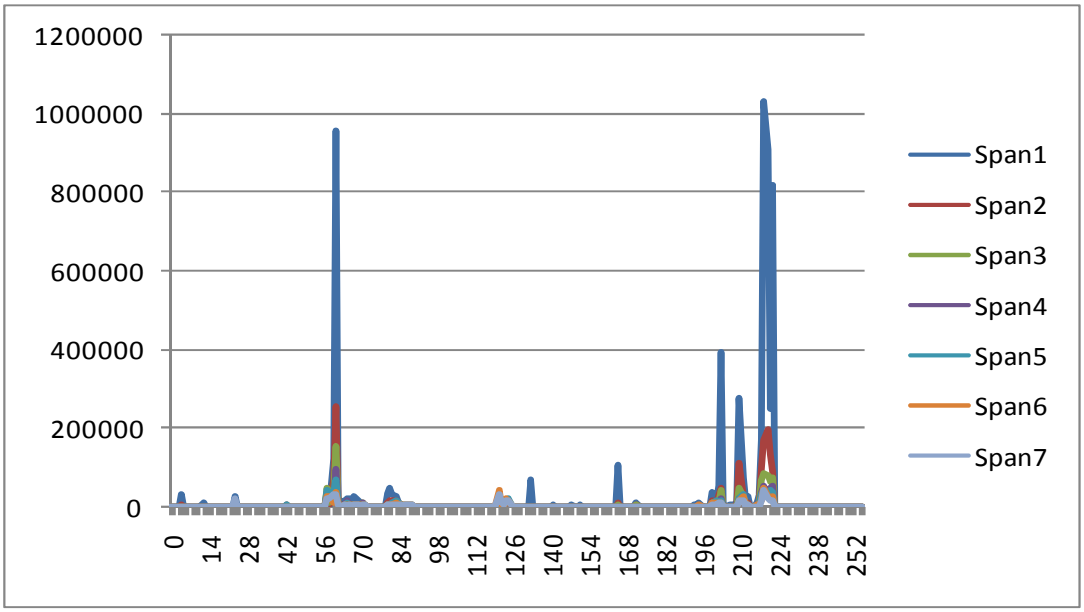

図 3：各期間の第一オクテットのヒストグラム

図 3 より第一オクテットの 56～70、216～222 に集中してることがわかる。これより不正ホストは固まっていることがわかる。また、期間を過ぎるごとに累計が減っていることもわかる。

2) 日毎のパケット数

各期間の日毎のパケット数を次の図 4,5,6,7,8,9,10 に示す。

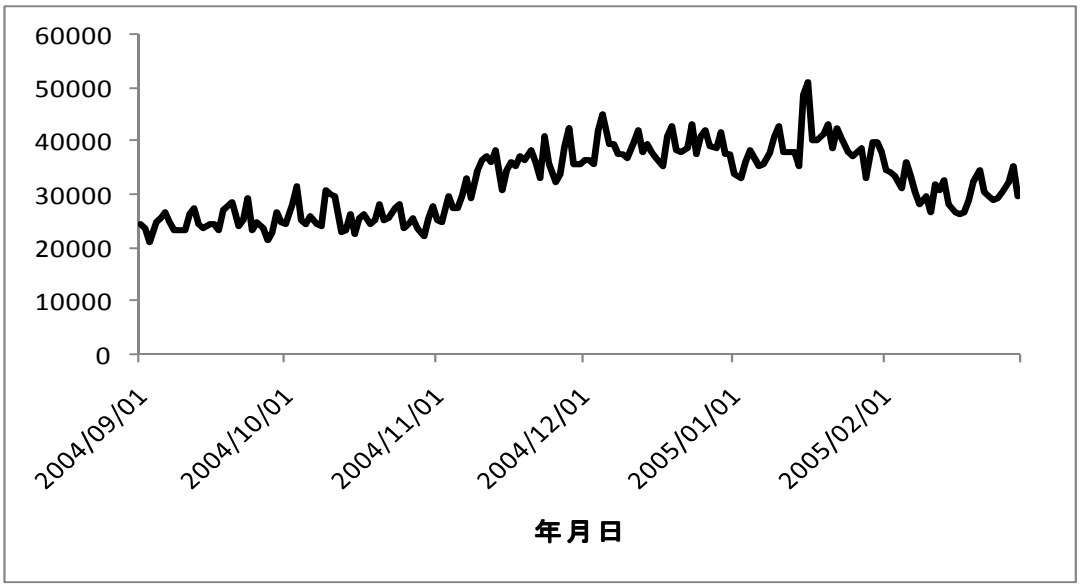

図 4:2004/9/1 から 2005/2/28 までの日毎のパケットの累計グラフ

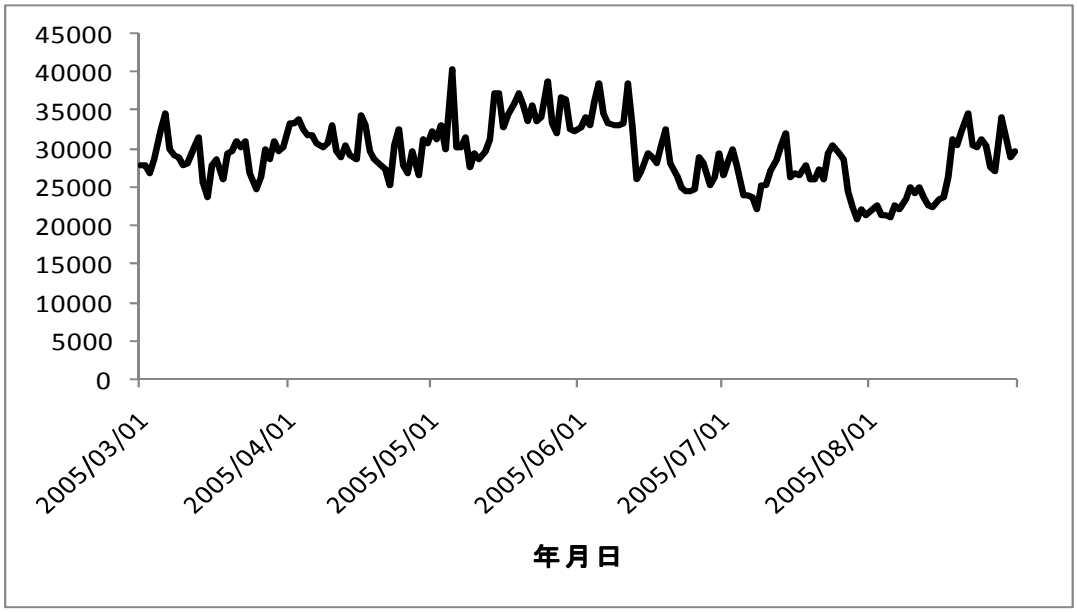

図 5:2005/3/1 から 2005/8/31 までの日毎のパケットの累計グラフ

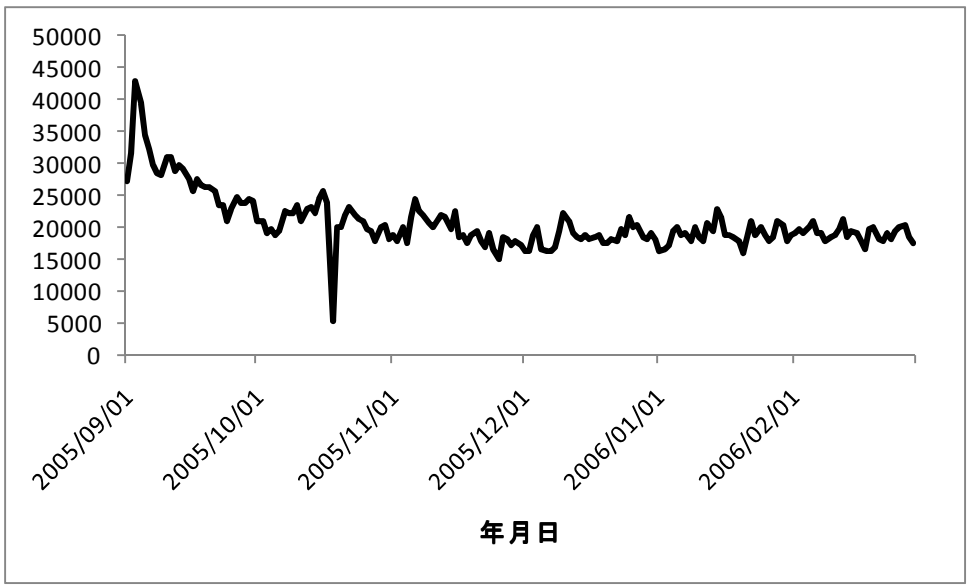

図 6:2005/9/1 から 2006/2/28 までの日毎のパケットの累計グラフ

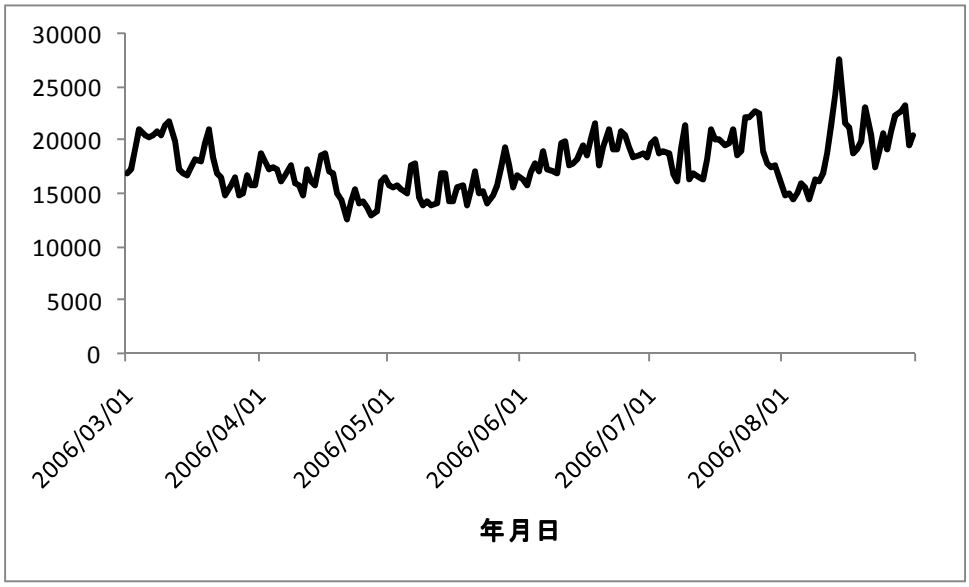

図 7:2006/3/1 から 2006/8/31 までの日毎のパケットの累計グラフ

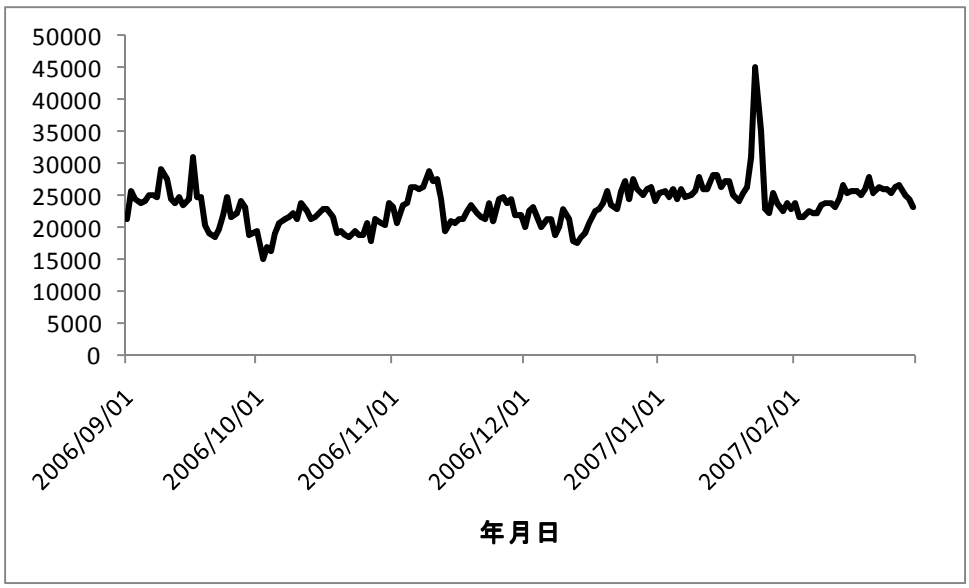

図 8:2006/9/1 から 2007/2/28 までの日毎のパケットの累計グラフ

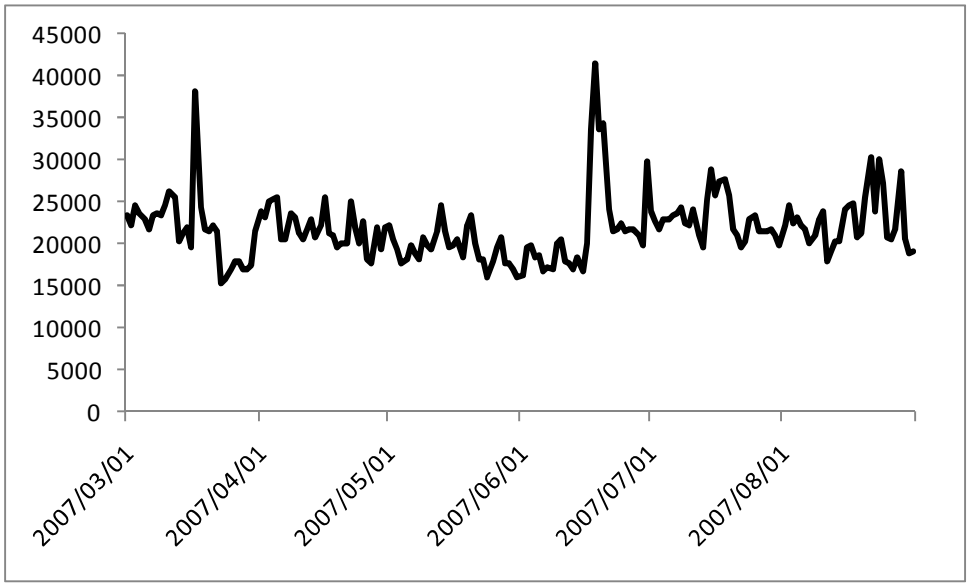

図 9 : 2007/3/1 から 2007/8/31 までの日毎のパケットの累計グラフ

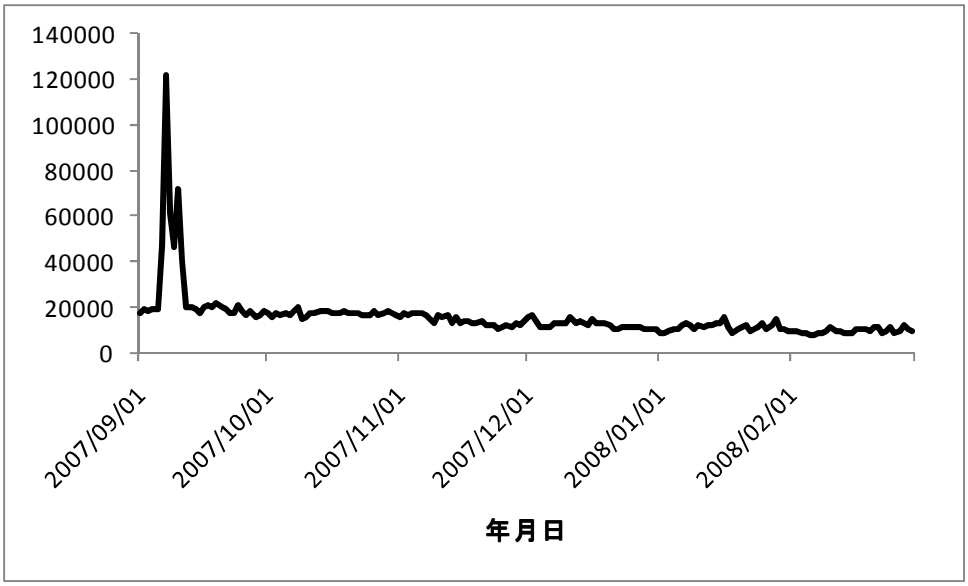

図 10：2007/9/1 から 2008/2/29 までの日毎のパケットの累計グラフ

各図から二つ特徴的なものが見えた。1 つめに、図 6 の期間で 2005/10/18 の日に急激な減少が見える。図 10 において 2007/9/7 に急激な増加が見えた。

3) 宛先ポート番号

各期間において、攻撃先、宛先ポートの順位を次の表 4 に示す。

表 4:宛先ポート順位

宛先port順位	Span1	Span2	Span3	Span4	Span5	Span6	Span7
1位	135	135	135	135	135	135	135
2位	445	445	445	445	445	445	4444
3位	137	139	139	1026	1026	1026	445
4位	139	1433	1026	139	139	139	1026
5位	1433	1026	80	80	1027	6346	138
6位	4899	80	1433	1433	1433	18120	520
7位	1025	137	1027	1027	8080	6646	139
8位	5554	4899	1434	1434	1434	1027	18120
9位	1434	1434	4899	4899	11234	520	1433
10位	2745	1027	137	137	80	8080	6646

表 4 より port135 への攻撃がずっと続いている。port135 は MS-RPC いうのは Windows 用のリモートのコンピュータ上のプログラムを呼び出すために開発された RPC（Remote Procedure Call）という機能を使うためのポートである。また port137,139,445,1433 なども攻撃先ポートで上位になっていることがわかる。それぞれ

NetBIOS, Direct Hosting SMB Service(ファイル共有), SQL Server 用のポートであった。

謝辞

本研究を遂行するにあたり，データの解析方法、論文の記述の仕方をご指導いただいた、東海大学電子情報学部情報メディア学科教授菊池浩明教授、東海大学院の小堀智弘様、共同研究者として、研究にお力を貸して頂いた、寺田真敏様、土居範久様、定点観測データ提供していただいたJPCERT/CC に厚い感謝を申し上げます。

参考文献

[1] 小堀，福野，菊池，寺田，土居,ISDAS 分散観測:ワームの平均寿命はいくらか？，コンピュータセキュリティシンポジウム(CSS2006)，情報処理学会，pp.519-524，2006.

[2] JPCERT/CC，ISDAS，(http://www.jpcert.or.jp/isdas)

[3] 福野,小堀,菊池,寺田,土居,インターネットの分散観測による不正侵入者のマクロ・ミクロ解析,情報処理研報，vol.2006，NO.81，2006-CSEC-34，pp.299-304，2006.

[4] 菊池，田中，福野，杉山，菊地，寺田，土居，ネットには不正ホストが何台いるか？，情報処理学会，コンピュータセキュリティシンポジウム(CSS2005)，pp.421-426，2005.